International Journal of Advance Research in Science and Engineering
Volume No. 13, Issue No. 12, December 2024
www.ijarse.com
IJARSE
ISSN 2319 - 8354

A Comprehension Analysis of Load-balancing Techniques & Classifications in Cloud Computing Environment

F. Niyasudeen[1], M. Mohan[2]

[1]Research Scholar, Department of Computer Science, SRM University, Delhi-NCR, Sonepat, Haryana, India

[2]Department of Computer Science and Engineering, SRM University, Delhi-NCR, Sonepat, Haryana, India

ABSTRACT

The recent phenomenal success of the Internet has opened access to computing resources, enabling the emergence of a novel computing concept known as cloud computing. Cloud computing users use platforms, infrastructure, and applications as services. Consequently, it is constantly expanding and changing to satisfy user demands. However, the expansion of services offered has brought about several difficulties and problems with load balancing in cloud computing. This study examines cloud computing platforms and implements load balancing techniques to enhance the overall efficiency and reduce response times of data centers for end-users, companies, and institutions. Load balancing provides optimized resources to enable various applications or software to operate smoothly for end users. Although cloud computing presents the modern IT sector with opportunities, there are still some challenges that need to be properly considered. We hope to deliver the best possible grasp of cloud computing. Our primary goal is to offer a comprehensive understanding of load balance in a cloud computing environment based on application design.

Keywords: *cloud computing, cloud types, Virtualization, Load balancing , challenges of load balancing.*

I. INTRODUCTION

At present, cloud computing is a widely adopted technology. This innovative approach is founded on the principles of distributed, parallel, and grid computing systems. It aims to provide users with access online using virtualized hardware and software infrastructure. Virtualization, networking, web services, distributed computing, and software are involved. It has also drawn significant attention from governments, businesses, and academics. Service providers create a cloud-computing environment. These companies offer services and infrastructure to end users.

Cloud computing, such as Google, Amazon, and Microsoft, has a greater impact on the IT sector. Several features have drawn attention to cloud computing. Cloud computing has received considerable attention from the commercial media. However, there is still disagreement over its precise definition and its relationship to grid computing. The fundamental components of cloud computing encompass the capacity to provide information technology infrastructure and applications that can be readily scaled up or down. These components form the foundation of cloud-based services. Internet-based cloud computing enables the use of operating systems and software applications through online platforms.

IJARSE
ISSN 2319 - 8354

Remote servers on the Internet are utilized by cloud computing technology to store, process, and manage data, eliminating the need for local servers or personal computers. Through internet connectivity, individuals can retrieve their files on any computer without the need to install additional software, allowing remote access to their resources. This computing process can be executed across multiple connected computers by leveraging the virtualization concept [1].

II. VIRTUALIZATION

Virtualization serves as the fundamental infrastructure for cloud-computing environments. The virtualization layer is positioned above the physical layer. Virtualization technology is extensively utilized in cloud computing data centers to minimize the quantity of hardware resources employed. This approach helps minimize capital expenses, energy consumption, and costs associated with cooling systems in cloud environments. A single physical server can support the simultaneous operation of several [virtual] servers, for example, through server consolidation. [2]. This idea can be implemented by aggregating unused resources into collective reservoirs and creating various categories of virtual machines. It performs various tasks concurrently and dynamically, by allocating resources.

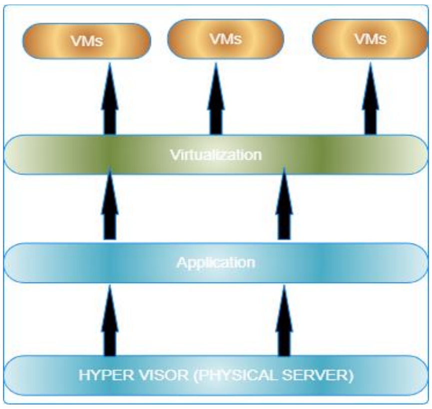

Fig.1. Layer of Virtualization

A. Virtualization Techniques

It is essential for individuals to possess a comprehensive understanding of fundamental virtualization methodologies, encompassing emulation, hypervisor, full virtualization, paravirtualization, and hardware-assisted virtualization. This program is referred to as a virtualization manager or a hypervisor. It acts as a mediator between hardware and operating systems [3].

III. FEATURES OF CLOUD COMPUTING

- A new computing paradigm is called as cloud computing.
- X-as-a-Service is used to provide applications and infrastructure resources such as hardware, storage, and system software. The foundation of cloud computing lies in its usage-based pricing models, where external providers or third-party clients offer these services.
- Virtualization and dynamic on-demand scalability are the two main cloud characteristics.
- A specified API or a web browser is used to access the cloud services [4].

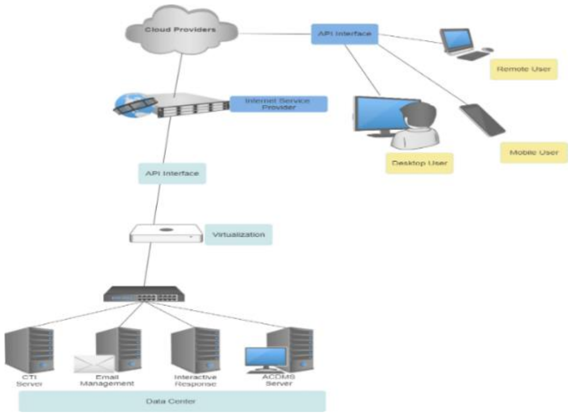

Fig. 2. Nature of Cloud Computing

A. Layers of Cloud Computing

Cloud Computing employs three architectural layers that are utilized by service providers for their clients. These layers form the foundation of Cloud Computing architecture. These are:

1) Infrastructure as a Service [IaaS]: This provides computational resources as a service, including processing power, data storage, and memory allocation. The Infrastructure as a Service (IaaS) model delivers customer-oriented solutions such as virtualized machines, storage systems, and disk image repositories. It also offers virtual infrastructure, raw block storage capabilities, file or object storage options, load balancing services, IP address allocation, firewall protection, and virtual local area networks. These resources are supplied on-demand from extensive pools housed within data center complexes.[4,5].

2) Platform as a Service [PaaS]: This framework involves cloud service providers offering a comprehensive computing platform as a service, typically encompassing a programming language within their operational environment, which includes operating systems, databases, and web servers. Software developers can create solutions on these cloud platforms without incurring the expenses associated with purchasing and managing the fundamental hardware and software components required for the execution environment. [4,6].

3) Software as a Service [SaaS]: The software-as-a-service approach is utilized by this system to deliver applications to end-users. Through an operational platform featuring an application and management interface, users can access the software and databases integral to the business model.[7].

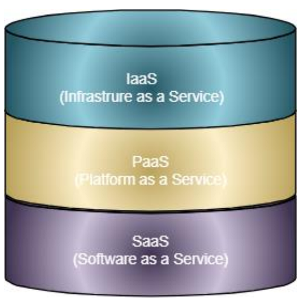

Fig. 3. Three layers of cloud computing

B. Deployment models of cloud computing

1. Public Cloud:

Users can interact with the cloud via web browser interfaces in this system. This helps reduce the operational costs of IT expenditure. Service providers manage all the types of resources and applications.

2. Private Cloud

This model separately deploys a cloud environment for an organization. Furthermore, private clouds offer exclusive cloud-based environments that are secure and dedicated to specific clients. Cloud refers to the on-premises cloud used to provide high-level control and give high security.

3. Hybrid Cloud

A hybrid approach combines elements of both public and private cloud systems. This model involves connecting a private cloud infrastructure to external cloud services. Hybrid cloud provides only few services to public cloud services in the organization and rest of the services as a private cloud.

4. Community Cloud

A community cloud is utilized when multiple organizations collaborate to build and utilize a shared cloud infrastructure service that aligns with their collective policies and needs.

C. Advantages and disadvantages of cloud computing

1) Advantages are,

- Easy, flexible, and Better Performance.
- Low cost compared to on-premises.
- Low-cost IT Infrastructure.
- Unlimited Storage Capacity based on end user needs.
- Universally, the data were available.
- Fewer trainings are sufficient for management.
- Availability of Latest Software for End Users.
- Increased Group Collaboration for improvement.

2) Disadvantages are,

- End-users need a constant Internet connection.
- End users do not work well with a slow Internet Connection.
- Lesser Security and Privacy for Data Management.
- Replication time and storage cost
- Inflexibility in accessing data and connectivity.

IV. LOAD BALANCING IN CLOUD COMPUTING ENVIRONMENT

In cloud computing, Load Balancing is employed to enhance the performance of Virtual Machines. The primary objective of load balancing techniques is to ensure an equitable and dynamic distribution of workload, while also maximizing resource utilization in the cloud environment. Effective workload management should result in a balanced system, which in turn leads to improved user engagement, increased satisfaction, and more efficient allocation of resources within the cloud infrastructure.

Load Balancing not only minimizes latency in data transmission but also prevents node overloading, ensuring that the quality of service in cloud data centers remains unaffected. [11]. In addition, it helps increase the response by distributing tasks equally for the application. End users can easily access software or web application interfaces with the help of load balancing. [11]. In a cloud environment, Load Balancers are utilized by service providers to efficiently handle online traffic by automatically allocating workloads across various servers and resources. This approach optimizes performance, minimizing response times and preventing system overload through the load balancer's effective distribution of tasks. Some load-balance algorithms target higher output, response time, and resource utilization [12].

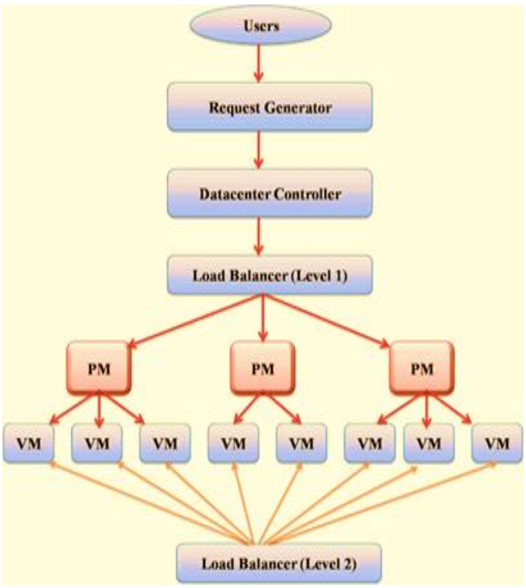

Fig. 4. Load balancing algorithm basic model

Load balancing is used to allocate small or large processes to nodes to enhance the overall performance of the system. Load distribution involves transferring workload from one node to another, enhancing resource efficiency and minimizing response times for concurrent tasks, thereby preventing delays or system failures.

Load Balancing has two parts that are important goals in a cloud environment. One is based on Resource Allocation and the other is Task Scheduling. First, VMs should be mapped to correct tasks such that there is no overload on the nodes. Following the allocation process, the second task is implemented. To fulfill the specific requirements delineated in the Service Level Agreement (SLA) document, task scheduling methodologies are implemented to optimize user requirements.[13]. The selection of a load-balancing algorithm is difficult because of its security, reliability, and throughput [14] [15].

A. Basic Criteria of Load balancing Algorithms

Three criteria were used to compare the load-balancing algorithms: Throughput, Execution Time, Processing Time. The calculation and estimation of the total number of jobs completed within a set timeframe utilize these three fundamental principles, excluding the time required for creating and terminating virtual machines. It is also used to measure the time taken to complete the process in virtual machines and the processing time in virtual machines [16,17]. The following load-balancing criteria select the algorithm for the scheduling tasks.

Step 1: Sending requests from the client’s side

Step 2: The cloud service provider distributes the task among data centers.

Step 3: Schedule the tasks in the data center by the service providers.

IJARSE
ISSN 2319 - 8354

B. Main goal of Load balancing[18,19]

For load balancing, there are several important goals. The first is the management and control of traffic on the server side. Second, the response time for end-user requests was reduced, and the ratio of resource usage was monitored. Third, it substantially improved its performance. Finally, the system firmness is maintained and checked, and the system's flexibility in the response time is increased.

C. Benefits of Load balancing

Cloud-based load balancing systems are evaluated using various criteria, including efficiency, speed of response, ability to scale, output rate, optimal use of resources, resilience to failures, duration of transfers, and related costs. Furthermore, we evaluate load-balancing strategies that prioritize energy efficiency by examining factors such as power usage and greenhouse gas production.[18,20].

TABLE I. TYPES OF LOAD BALANCING BENEFITS

Performance	Resource Utilization	Scalability	Response Time
Evaluating a system's efficiency involves assessing its performance. For instance, this can be achieved by minimizing response times while maintaining acceptable latency levels.	Resource Utilization is used to check the utilization of resources. It should be optimized for efficient load balancing.	An algorithm's capacity to distribute workload evenly across a system containing any finite quantity of nodes. This performance indicator requires enhancement.	The amount of time taken to respond by a particular load-balancing algorithm for a distributed system.

D. Classification of Load balancing

There are two main classifications for load-balancing algorithms: static and dynamic.[21]. A static load-balancing algorithm disregards a node's previous condition or performance when distributing the workload. In contrast, a dynamic load-balancing algorithm examines a node's prior status before assigning tasks. The biggest benefit of dynamic load balancing is that a single node failure does not result in system failure; however, it has a minor impact on how well the system operates. This approach encompasses various strategies, including transfer, information, selection, and the choice of location policy. The dynamic load balancer works based on four types of policies for keeping track of updated information: transfer policy, selection policy, location policy, and information policy [17].

a) Round Robin and Randomized [RRRA]: Static

- The workload is evenly distributed among all processors by dividing processes in a way that ensures equal distribution across the processing units.
- Each processor recorded the order in which the tasks were assigned.
- Handling user requests in a circular manner.

b) Central Manager: Static

- For every new process, the host must be chosen by a central processor.

- The processor with the lowest capacity is identified by the overall workload established during the initial setup phase.

c) Min-Min: Static

- All the jobs were based on the shortest possible completion.
- The minimum number of iterations was used to calculate the minimum value.
- This work is based on this minimal time.

d) Biased Random sampling: Dynamic

In this system, each node's vertex is represented by a server, while the node's available free resources are indicated by its degree. The allocation is determined based on the degree value. Work is assigned to each node if it has more than one degree. The degree of a node increases once the work is finished and decreases by one when the work is assigned to it.

e) Active Clustering: Dynamic

The core concept of this algorithm involves clustering similar nodes and working with these clustered units. By consolidating nodes, this resource can more effectively improve performance and output.

f) Ant Colony and Complex network Theory [ACCLB]: Dynamic

Effective load balancing is made possible by a complex network's small world and size freedom [22,23].

V. PROBLEMS OF LOAD BALANCING IN THE CLOUD

More in-depth theoretical and practical research in the data and computation domains is increasingly focused on cloud computing technology. Cloud computing research faces numerous challenges, with load balancing being a key issue that demands particular focus. Furthermore, several other concerns require equal consideration to optimize cloud resource utilization, including VM migration, VM security, meeting user QoS expectations, and resource management. The following section outlines some of the primary load-balancing challenges encountered in this field.:

a) Geographically distributed nodes: Geographically dispersed data centers are used in cloud computing. Techniques for load balancing are limited in scope and ignore variables such as communication and network latency, the separation between dispersed computing nodes, and the distance between users and resources. These algorithms are unsuitable for this environment, making nodes situated at great distances a challenge. Consequently, it is important to consider developing algorithms for load balancing that are effective for geographically distant nodes.

b) Virtual Machine Migration: Multiple virtual machines (VMs) can be created on a single virtualized machine that is virtualization. These virtual machines are self-contained and have various setups. When a physical machine becomes overloaded, specific virtual machines (VMs) need to employ a VM load-balancing strategy to relocate to a different site.[26].

c) Storage Management: Cloud storage represents an innovative alternative to conventional storage systems, which typically involve substantial hardware expenses and management requirements. This modern approach allows users to store diverse types of data without encountering access difficulties or cloud-related issues. As the volume of data in cloud storage continues to expand rapidly, it becomes necessary to implement data replication strategies to ensure consistent and reliable access. Full data replication strategies become inefficient when identical data storage policies are implemented across replication points. Although partial replication may be adequate, it can complicate load-balancing strategies and cause problems with dataset availability. Therefore, it is necessary to develop an efficient load-balancing technique for a partial replication system that considers both the distribution of the application and its associated data. [27].

d) Load-Balancer Scalability: The flexibility and adaptability of cloud services enable users to access resources as needed, allowing for rapid expansion or reduction of usage. To efficiently manage these fluctuations, implementing a robust load balancer is crucial for addressing unexpected changes in resource requirements, including storage capacity, computational power, network configuration, and other variables.

e) Algorithm Complexity: Cloud computing algorithms should be designed for ease of use and simplicity. Cloud system efficiency and performance are reduced by a complicated algorithm.

CONCLUSION AND FUTURE WORK

New cloud computing technology has received considerable research attention. Research attention has been drawn to load balancing, a primary task in virtual machines (VMs) and a major challenge in cloud computing. It is also a major challenge to comprehend and use these programs. Review different cloud computing types, applications, architectures, features, applications, deployment models, etc.. This study presents an extensive analysis of the ideas behind Cloud Computing technology, a variety of cloud service types, and difficulties with load balancing in cloud computing environments. When cloud computing advances, more academics, businesses, organizations, and individuals find it an affordable and practical option to shift their information processing to the cloud. However, more research is still required to prevent problems and delay responses in cloud computing, as well as to provide better performance and improve the quality of services (QoS) in the IT industry.

REFERENCES

[1] Fatima, Syeda and Abdul Sattar, Shahid and Khan, Naseer and Adil, Syed. [2019]. cloud computing and load balancing. international journal of advanced research in engineering & technology. volume 10, issue 2, march-april 2019, pp. 189-209. 10. 10.34218/ijaret.10.2.2019.019. t http://www.iaeme.com/ijaret/issues.as

[2] Devi N.Nirmala, Kumar S.Vengatesh, A Study on Virtualization in Cloud Computing, International Journal of Communication and Networking System Volume: 08 Issue: 02 December 2019 Page No.70-75 ISSN: 2278-2427.

[3] Bhupender, “An overview of types of virtualization in cloud computing”, Jun 2016, Cloud , VPS,https://www.znetlive.com/blog/virtualization-in-cloud-computing/.

[4] K. Stanoevska-Slabeva et al. [eds.], Grid and Cloud Computing: A Business Perspective on Technology and Applications, DOI 10.1007/978-3-642-05193-7_4.

[5] Ayesha Malik, Muhammad Mohsin Nazir, “Security Framework for Cloud Computing Environment: A Review”, Journal of Emerging Trends in computing and information Sciences, March 2012.

[6] Qi Zhang et al.”Cloud Computing: state-of-the-art and research challenges”,The Brazilian Computer Society 2010, Apr 2010.

[7] Rydhm Beri and Veerawali Behal, Cloud Computing: A Survey on Cloud Computing, International Journal of Computer Applications [0975 – 8887] Volume 111 – No 16, February 2015,19-22.

[8] Yashpalsinh Jadeja and Kirit Modi, Cloud Computing - Concepts, Architecture and Challenges, Intenational Conference on Computing, Electronics and Electrical Technologies, 978-1-4673-0210-4112/$31.00 ©2012 IEEE,877-880.

[9] Ramya Mohanakrishnan, A community cloud is a cloud infrastructure in which multiple organizations share resources and services based on common requirements. https://www.spiceworks.com/tech/cloud/articles/what-is-community-cloud/ 5th August 2021.

[10] Taqwan Thamrin , Marzuki , Reni Nursyanti , Andala Rama Putra, 1 st International Conference on Engineering and Technology Development ,75-79,ISSN 2301-6590 [ICETD 2012].

[11] Dalia Abdulkareem Shafiq, N.Z. Jhanjhi, Azween Abdullah,Load balancing techniques in cloud computing environment: A review,Journal of King Saud University - Computer and Information Sciences,Volume 34, Issue 7,2022,Pages 3910-3933,ISSN 1319-1578, https://doi.org/10.1016/j.jksuci.2021.02.007.[https://www.sciencedirect.com/science/article/pii/S131915782100046X].

[12] Ananth Rao, Karthik Lakshminarayanan, Sonesh Surana, Richard Karp, and Ion Stoica. "Load Balancing in Structured P2P Systems." In Peer-to-Peer Systems II, pp. 68-79. Springer Berlin Heidelberg, 2003.

[13] D.A. Shafiq, N. Jhanjhi, A. Abdullah,Proposing A Load Balancing Algorithm for the Optimization of Cloud Computing Applications MACS 2019–13th Int Conf. Math. Actuar. Sci. Comput. Sci. Stat. Proc. [2019], 10.1109/MACS48846.2019.9024785

[14] H. Shoja, H. Nahid, and R. Azizi, “A comparative survey on load balancing algorithms in cloud computing,” 5th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT, 2014, no. February, pp. 3–8, 2014, doi: 10.1109/ICCCNT.2014.6963138.

[15] K. R. R. Babu and P. Samuel, “Innovations in Bio-Inspired Computing and Applications,” Adv. Intell. Syst. Comput., vol. 424, no. Ibica, pp. 139–149, 2016, doi: 10.1007/978-3-319-28031-8.

[16] N. Peter, “FOG Computing and Its Real Time Applications,” Int. J. Emerg. Technol. Adv. Eng., vol. 9001, no. 6, p. 266, 2015, [Online]. Available: www.ijetae.com.

[17] N. Ajith Singh, M. Hemalatha, ―An Approach on Semi Distributed Load Balancing Algorithm for Cloud Computing Systems‖ International Journal of Computer Applications Vol-56 No.12 2012.

[18] Yelchuri Venkata Sai Harsha , Nagaraj G Cholli, Load Balancing in Cloud Computing, International Journal of Advanced Research in Computer and Communication Engineering Vol. 10, Issue 5, May 2021, DOI 10.17148/IJARCCE.2021.105146. ISSN [Online] 2278-1021.705-709.

[19] Ram Prasad Padhy. "Load Balancing in Cloud Computing Systems." PhD diss., National Institute of Technology, Rourkela, 2011.

[20] Kansal, Nidhi Jain, and Inderveer Chana. "Cloud Load Balancing Techniques: A Step Towards Green Computing." IJCSI International Journal of Computer Science Issues 9, No. 1 [2012]: 238-246.

[21] Afzal, S., Kavitha, G. Load balancing in cloud computing – A hierarchical taxonomical classification. *J Cloud Comp* **8**, 22 [2019]. https://doi.org/10.1186/s13677-019-0146-7.

IJARSE
ISSN 2319 - 8354

[22] Shah, Jaimeel & Kotecha, Ketan & Pandya, Sharnil & Choksi, D.B. & Joshi, Narayan. [2017]. Load balancing in cloud computing: Methodological survey on different types of algorithm. 100-107. 10.1109/ICOEI.2017.8300865.

[23] Reena Panwar , Prof. Dr. Bhawna Mallick ”Load Balancing in Cloud Computing Using Dynamic Load Management Algorithm”IEEE 2015.

[24] Sandeep Negi, “An Improved Round Robin Approach using Dynamic Time Quantum for Improving Average Waiting Time” International Journal of Computer Applications Vol. 69 : No 14, May 2013..

[25] Shang-Liang Chen , Yun-Yao Chen ∗, Suang-Hong Kuo “ CLB: A novel load balancing architecture and algorithm for cloud services.

[26] Khan, Nooreen Fatima, and M. Mohan. "Cloud security using self-acting spontaneous honeypots." *International Journal of Engineering & Technology* 7.2.8 (2018): 243-247.

[27] Khan, Nooreen Fatima, and M. M. Mohan. "Honey pot as a service in cloud." *International Journal of Pure and Applied Mathematics* 118.20 (2018): 2883-2888.